

RESEARCH ARTICLE

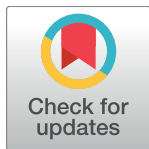
# Circulating miRNAs, isomiRs and small RNA clusters in human plasma and breast milk

Mercedes Rubio<sup>1</sup>✉, Mariona Bustamante<sup>2,3,4,5</sup>✉, Carles Hernandez-Ferrer<sup>2,4,5</sup>, Dietmar Fernandez-Orth<sup>2,4,5</sup>, Lorena Pantano<sup>6</sup>, Yaris Sarria<sup>7</sup>, Maria Piqué-Borras<sup>8</sup>, Kilian Vellve<sup>9</sup>, Silvia Agramunt<sup>9</sup>, Ramon Carreras<sup>9,10</sup>, Xavier Estivill<sup>11,12</sup>, Juan R. Gonzalez<sup>2,4,5\*</sup>, Alfredo Mayor<sup>1,13\*</sup>

**1** ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clínic—Universitat de Barcelona, Barcelona, Spain, **2** ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain, **3** Genomics and Disease Group, Bioinformatics and Genomics Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, **4** Universitat Pompeu Fabra (UPF), Barcelona, Spain, **5** CIBER Epidemiología y Salud Pública, Barcelona, Spain, **6** Harvard TH Chan School of Public Health, Boston, MA, United States of America, **7** Microarray Analysis Service, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain, **8** Laboratory of Childhood Leukemia, Department of Biomedicine, University of Basel and Basel University Children's Hospital, Hebelestrasse, Basel, Switzerland, **9** Obstetrics and Gynaecology Department, Hospital del Mar, Parc de Salut Mar, Barcelona, Spain, **10** Pediatrics, Obstetrics and Gynecology and Preventive Medicine Department, Universitat Autònoma de Barcelona, Barcelona, Spain, **11** Genetics of Child and Woman's Health Group, Research Department, Sidra Medical and Research Center, Doha, Qatar, **12** Genetics Unit, Dexeus Woman's Health, Barcelona, Spain, **13** Centro de Investigação em Saúde da Manhiça (CISM), Maputo, Mozambique

✉ These authors contributed equally to this work.

\* [juanr.gonzalez@isglobal.org](mailto:juanr.gonzalez@isglobal.org) (JRG); [alfredo.mayor@isglobal.org](mailto:alfredo.mayor@isglobal.org) (AM)



## OPEN ACCESS

**Citation:** Rubio M, Bustamante M, Hernandez-Ferrer C, Fernandez-Orth D, Pantano L, Sarria Y, et al. (2018) Circulating miRNAs, isomiRs and small RNA clusters in human plasma and breast milk. PLoS ONE 13(3): e0193527. <https://doi.org/10.1371/journal.pone.0193527>

**Editor:** George Calin, University of Texas MD Anderson Cancer Center, UNITED STATES

**Received:** August 2, 2017

**Accepted:** February 13, 2018

**Published:** March 5, 2018

**Copyright:** © 2018 Rubio et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Specific sequencing data for each participant are available from the GEO database (accession number GSE107524).

**Funding:** This work was supported by ISCIII—Subdirección General de Evaluación Plan Nacional I+D+I 2013–2016 (grant PI13/01478 cofunded by the Fondo Europeo de Desarrollo Regional [FEDER] and CES10/021-I3SNS to AM), the Departament d'Universitats i Recerca de la Generalitat de

## Abstract

Circulating small RNAs, including miRNAs but also isomiRs and other RNA species, have the potential to be used as non-invasive biomarkers for communicable and non-communicable diseases. This study aims to characterize and compare small RNA profiles in human biofluids. For this purpose, RNA was extracted from plasma and breast milk samples from 15 healthy postpartum mothers. Small RNA libraries were prepared with the NEBNext® small RNA library preparation kit and sequenced in an Illumina HiSeq2000 platform. miRNAs, isomiRs and clusters of small RNAs were annotated using seqBuster/seqCluster framework in 5 plasma and 10 milk samples that passed the initial quality control. The RNA yield was 81 ng/mL [standard deviation (SD): 41] and 3985 ng/mL (SD: 3767) for plasma and breast milk, respectively. Mean number of good quality reads was 4.04 million (M) (40.01% of the reads) in plasma and 12.5M (89.6%) in breast milk. One thousand one hundred eighty two miRNAs, 12,084 isomiRs and 1,053 small RNA clusters that included piwi-interfering RNAs (piRNAs), tRNAs, small nucleolar RNAs (snoRNA) and small nuclear RNAs (snRNAs) were detected. Samples grouped by biofluid, with 308 miRNAs, 1,790 isomiRs and 778 small RNA clusters differentially detected. In summary, plasma and milk showed a different small RNA profile. In both, miRNAs, piRNAs, tRNAs, snRNAs, and snoRNAs were identified, confirming the presence of non-miRNA species in plasma, and describing them for the first time in milk.

Catalunya (Agencia de Gestión de Ayudas Universitarias y de Investigación (AGAUR); grant 2014SGR263 to AM) and the CERCA Institutes Integration Program (SUMA 2013 to JRG, AM; Secretaría de Universidades e Investigación del Departamento de Economía y Conocimiento de la Generalidad de Cataluña, AGAUR and Fundació Institució dels Centres de Recerca de Catalunya). ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Discovery of novel biomarkers for early detection and improved prognosis in non-communicable and infectious diseases is a research priority. A good biomarker should be suitable to diagnose the disease, predict its progression or regression, or to monitor the outcome after treatment and, ideally, should be easily obtained with minimum invasion [1].

Promising classes of molecular biomarkers are small RNAs, and special attention has been given to microRNAs (miRNAs) [2]. miRNAs are short (21–23 nt), single-stranded, non-coding RNAs that regulate gene expression and play essential roles in fundamental biological processes. Recent studies have shown the involvement of miRNAs in various aspects of major chronic diseases including bronchial asthma and diabetes, as well as in regulation and induction of senescence, brain function and cancer [3–8]. Changes in miRNA expression can be triggered by environmental factors such as exposure to pesticides, heavy metals, air pollution, bisphenol A and cigarette smoking [9, 10]. The content of miRNAs is also influenced by host-pathogen interactions, as has been shown for bacteria, viruses and apicomplexan parasites [11].

miRNAs can be actively secreted to circulation packaged in microvesicles, exosomes, or associated to RNA-binding proteins, such as Ago2, but cell death or apoptosis can also release miRNAs into the circulation [9]. Their resistance to pH variations, their stability during long term storage and repeated freeze-thaw cycles [12], and the fact that they can be detected in circulation but released by distant organs constitute them as potential non-invasive biomarkers to monitor the body's pathophysiological status. However, besides some particular cases where circulating miRNAs have been tested in clinical trials [13, 14], there is lack of reproducibility between studies, highlighting the need of a standardization of laboratory and bioinformatic methods [15].

Next generation sequencing (NGS) allows not only the quantification of known miRNAs, but also the identification and quantification of novel miRNAs, isomiRs (miRNA variants) [16], and other small RNA species that can be functionally relevant in diseases and therefore used as potential disease biomarkers [17, 18]. However, only few studies have investigated circulating non-miRNA small RNAs, focusing on plasma [19, 20], urine and saliva [21]. Although there is an extensive literature on the potential effects of breast milk miRNAs on offspring development [22–24], none of these previous studies has evaluated non-miRNA small RNAs in human milk.

The vast majority of small RNA bioinformatic tools are designed for the analysis of miRNAs. They usually discard multi-mapping reads derived from non-coding RNAs with duplication events in the genome, or alternatively, they count them several times or split the reads in all the genomic positions, which bias the estimations. Recently, Pantano et al. developed SeqCluster, a method for the accurate quantification of small RNAs, based on the estimation of clusters of highly similar sequences irrespective of their genomic annotation [25, 26].

This study aims to characterize miRNAs, isomiRs and small RNA clusters in plasma and milk as well as compare their profiles. For this purpose, miRNAs, isomiRs and small RNA clusters were analysed in samples collected from 15 healthy women at 24–72h post-delivery. In order to facilitate future studies in the field, detailed information on sample collection, processing and bioinformatics analysis is given.

## Materials and methods

### Study population

Fifteen Spanish healthy volunteers were enrolled in the study. Exclusion criteria were as follows: HIV, Hepatitis B or Hepatitis C positivity; severe preeclampsia; gestational diabetes and

preterm delivery (<37 weeks of gestation). Basic demographics of the women and data about gestation and delivery were collected. Peripheral blood was collected at 0–48 h post-partum for 15 participants; and breast milk was collected at 48–72 h post-partum for 10 of the participants. Participants signed an informed consent and protocols were approved by the Hospital del Mar and Hospital Clinic (Barcelona, Spain) ethical committees.

## Collection and pre-processing of samples

**Plasma.** Twelve mL of peripheral blood were collected in VACUETTE® EDTA Tubes (K3EDTA vacutainers) (Greiner Bio-One, Cat No.: 456038). EDTA blood tubes were gently mixed by inverting 10 times and stored at 4°C upright until centrifugation. Plasma separation was performed by centrifugation of blood samples in horizontal rotor for 10 minutes at 2000  $\times$  g at 4°C. A second centrifugation at 16,000  $\times$  g for 15 minutes at 4°C, was performed to eliminate any cell debris. Plasma was transferred to a new RNase-free cryotube avoiding the collection of any cell debris and stored at -80°C.

Plasma haemolysis was evaluated by spectrophotometry using a NanoDrop ND-2000 equipment (Thermo Fisher Scientific). Absorbance at 414 nm of 0.2 was set as cut-off haemolysis as described before [27].

**Milk.** Fifteen mL of breast milk were extracted using an electric pump (Medela, Cat No.:008.0176) following manufacturer's instructions and transferred to a 100 mL sterile RNase-free recipient. Milk was kept in ice and vortexed gently before centrifugation at 2000  $\times$  g for 10 minutes at 4°C to remove milk fat globules, cells, and large debris. Immediately after carefully transferring, supernatant was carefully transferred into new 2 mL RNase-free tubes, a second centrifugation at 16,000  $\times$  g for 15 minutes was performed at 4°C to remove residual fat, cell debris and the casein fraction. Skim milk, avoiding fat or cells, was transferred into 2.5 mL cryotubes and stored at -80°C.

## RNA extraction, quantification and quality control

Small RNAs were extracted using the miRNeasy Serum/Plasma kit (Qiagen, Cat No.: 217184) with minor modifications. RNA extraction was performed from at least 3 mL of plasma and milk. One mL of biofluid was filtered per each column and washed two times with ethanol 80% before elution. Each sample was measured for RNA quantity using Fluorometric quantification (Qubit™ 2.0 Fluorometer, Thermo Fisher Scientific) with the Qubit miRNA assay (Thermo Fisher Scientific) and for integrity and quality using two different methods: spectrophotometer (NanoDrop™ 1000, Thermo Scientific) and electrophoresis and flowcytometry (Agilent RNA 6000 pico chip and Agilent 2100 Bioanalyzer, Agilent Technologies).

## Library preparation and sequencing

Libraries were prepared using the NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) (NEB, Cat No.: E7330) following manufacturer instructions and using 150ng of RNA obtained from plasma samples and 500 ng from milk. Library size selection was done using acrylamide gels. Quality and concentration of cDNA libraries was checked, and finally groups of 13–20 samples were pooled at the same concentration before sequencing in an Illumina HiSeq2000 Sequencing System (50 nt, single read).

## Bioinformatic and statistical analyses

**Quality control.** The bioinformatic pipeline consisted of several steps (S1 Fig). An initial quality control (QC) was conducted using FASTX-Toolkit and FastQ Screen. After adaptor

removal, reads with the following features were removed: 1) Reads <18nt, 2) Mean PHRED scores < 30, and 3) Low complexity reads based on mean score of the read. Then, QCed reads were mapped to miRNAs, and miRNA complexity was estimated as the number of miRNA genes that are observed as a function of the number of miRNA reads.

**Identification and quantification of small RNAs.** Sequences fulfilling QC were used as the input for the seqBuster/seqCluster tool that retrieves three matrices of counts: i) miRNA, ii) isomiRs, and iii) small RNA clusters [25, 28]. In order to detect miRNAs and isomiRs, reads were mapped to hairpins released from miRBasev21 using miraligner allowing one mismatch [28]. Sequences mapping ambiguously to several positions of the genome were filtered out. Only isomiRs with 3' or 5' trimming or 3' addition modifications are kept in the result files. isomiRs with one nucleotide substitution are not reported due to the difficulty of linking them unambiguously to a particular miRNA. IsomiR naming followed the miRTop project's recommendations (<http://mirtop.github.io/>). IsomiR names can be merged with isomiR sequences using a GFF3 format file provided on GEO. In parallel, sequences were mapped to hs37d5 using bowtie [29], and hotspots (sets of overlapping sequences according to their position in the genome) were detected. Hotspots sharing any sequence were grouped in primary clusters. Then, a recursively heuristic algorithm based on reduction and cluster correction was applied to the ones sharing over 60% of the sequences. These were annotated to different RNA species using miRBase v20 [30], refGene (RefSeq genes), wgRna (CD and H/ACA Box snoRNAs and miRNAs from Weber and Griffiths-Jones), rnsk (Repeating elements created using Repeat-Masker) and tRNA from University of California Santa Cruz (UCSC) genome browser (hg19) [31], and piR\_hg19\_v1.0 from the regulatoryrna database ([www.regulatoryrna.org](http://www.regulatoryrna.org)) [32]. Clusters can be considered as unique units of transcription, regardless of their annotation to one or multiple positions in the genome (irrespective of their genomic origin).

Raw FASTQ files, quality controlled unnormalized counts, metadata and a GFF3 format file with isomiR annotation can be downloaded from Gene Expression Omnibus (GEO): GSE107524.

**Normalization and differential expression.** Normalization and differential expression was performed with the R package DESeq2 v.1.10.1 (R version 3.3.2) [33]. DESeq2 performs an internal normalization where geometric mean is calculated for each gene across all samples (scaling factor method) The impact of main technical and demographic/biological variables on normalized counts was explored through Principal Component Analysis (PCA) and dendrograms implemented within the FactoMineR R package [34]. Differential expression was assessed with negative binomial generalized linear models adjusting for multiple testing with the False Discovery Rate (FDR) method [35]. All models considered biofluid as the independent variable and were adjusted for subject as a fixed effect. In addition, a sensitivity analysis considering only paired samples was performed and results did not change substantially.

Milk samples presented systematically higher sequencing depth than plasma samples. As sensitivity analyses, we repeated the miRNA differential analyses with different filtering approaches. Our main analysis used DESeq2 to filter out features having little chance of showing significant evidence as they contain outliers or low mean normalized counts. Second, to avoid comparison of miRNAs not detected due to the lower sequencing depth in plasma samples, an initial filtering of miRNAs present in <80% of the samples and with <10 reads was done. Finally, ten random milk and plasma subsamples of 0.2M miRNA reads each (the minimum in plasma samples) were obtained and analysed. P values of each of the 10 differential analysis were combined using simes test [36], from mppa v.1.0 (eprint arXiv:1408.3845) in R (3.3.2).

## Results

### Study population and sample collection

Fifteen healthy volunteer mothers were enrolled in the study during 24–72h post-delivery. Main characteristics of the mothers are found in Table 1. Mean age of the participants was 32.9 years. Forty percent of them were primiparous and only one of them smoked during pregnancy. One third of the participants were born in Spain, one third in Asia and the rest were from Morocco and South America. Plasma samples were collected for all 15 women, while milk was available for 10 of them.

### RNA extraction and quality control

The average amount of RNA extracted from plasma and milk samples was 81.14 ng/mL (standard deviation, SD 40.92) and 3,984.85 ng/mL (SD 3,766.97), respectively (S1 Table). In the Bioanalyzer plot, all biofluids showed a peak between 25 to 200 nucleotides which corresponded to the expected size for small RNAs (S2 Fig). No peaks were detected at 18S and 28S in plasma, whereas a peak was observed in half of the milk samples, suggesting potential contamination with cellular RNA.

### Quality control

The mean number of reads obtained for each biofluid was similar: 13.96 million (M) in plasma and 13.87 M in milk. However, the average proportion of good quality reads was 89.58% in milk, and only 40.01% in plasma (Table 2 and S2 Table). Three point ninety seven percent and 35.72% of the raw reads mapped to miRNAs in plasma and milk, respectively. Only samples with >2M of good quality reads of which >0.2 M mapped to miRNAs were considered for further analysis (10 milk and 5 plasma samples).

### miRNAs

Total number of miRNAs detected with >1 count were 1,182: 1,002 in milk [456 with a mean abundance of >10 reads per million (RPM)] and 824 in plasma [271 with a mean abundance of >10 RPM]. Ten miRNAs comprised >70% of the expression in each biofluid (Fig 1, S3 Table). miRNA complexity was quite heterogeneous among samples, even among samples from the same biofluid, and tended to be higher in milk than in plasma (S3 Fig). Samples clustered by biofluid and not by subject (Fig 2). Data did not cluster by technical factors (time to

**Table 1. Main characteristics of study participants.**

Variable	Summary
Age (years); mean (range)	32.93 (23–40)
Parity; n (%)	
Primigravidae	6 (40%)
Multigravidae	9 (60%)
Region of origin; n (%)	
Spain	5 (33.3%)
Asia	5 (33.3%)
Morocco	2 (13.3%)
South America	3 (20.1%)
Tobacco smoking during pregnancy; n (%)	
Yes	1 (6.7%)
No	14 (93.3%)

<https://doi.org/10.1371/journal.pone.0193527.t001>

**Table 2. Quality control and mapped reads to miRNAs [mean and (SD)], by biofluid.**

		Plasma	Milk
		(N = 15)	(N = 10)
Total reads; millions (SD)		13.96 (16.71)	13.87 (8.23)
Quality filtered; millions (SD)		9.82e-3 (9.75e-3)	6.33e-3 (1.98e-3)
Complexity filtered; millions (SD)		1.40e-4 (1.32e-4)	1.39e-4 (7.26e-5)
Size filtered; millions (SD)		9.92 (11.40)	1.36 (0.86)
Good quality reads <sup>a</sup>			
	Millions (SD)	4.04 (6.16)	12.52 (7.74)
	% raw reads (SD)	40.01 (23.05)	89.58 (5.42)
miRNA			
	Millions (SD)	0.27 (0.44)	5.42 (4.88)
	% raw reads (SD)	3.97 (4.36)	35.72 (12.40)
	% QCed reads (SD)	8.92 (6.85)	40.02 (13.43)
N samples tested <sup>b</sup>		5	10

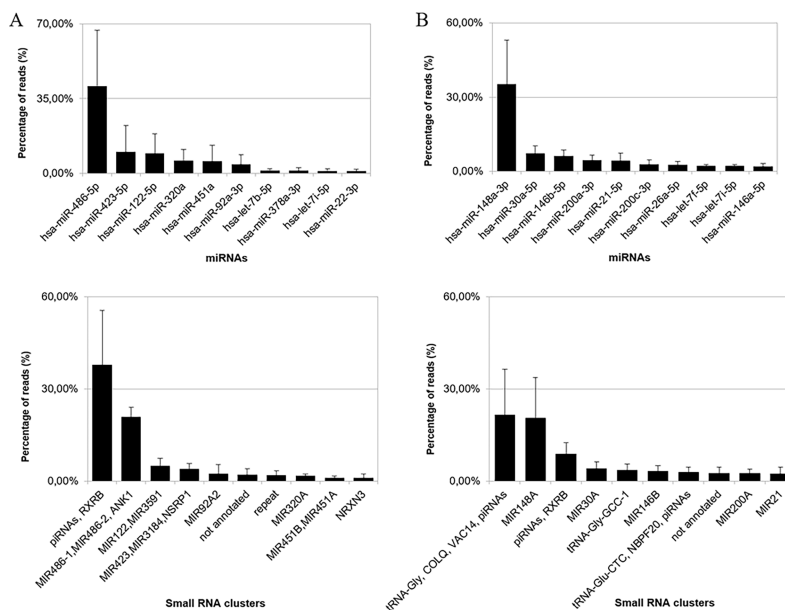
<sup>a</sup> Reads after filtering low quality, low complexity and size <18 nt.

<sup>b</sup> Number of samples tested in the downstream differential analysis of miRNAs, isomiRs and small RNA clusters (>0.2M reads in miRNAs).

<https://doi.org/10.1371/journal.pone.0193527.t002>

storage, RNA extraction date, library preparation date, sequencing run and lane) or demographic/biological variables (age, region of origin, and parity) (S4 Fig).

At 5% FDR, 308 miRNAs showed different levels between biofluids [126 (40.91%) milk > plasma] (Table 3 and S4 Table). Among the 112 miRNAs found in >80% of the samples and with >10 counts, 87 of them showed statistically significant differences between biofluids (66 milk > plasma). Seventy-four out of the 87 miRNAs overlapped with the ones detected in the main analysis (S7 Fig).



**Fig 1. Relative abundance of miRNA and small RNA clusters.** (A, C) Plasma, (B, D) milk. The x-axis represents miRNAs (A, B) or small RNA clusters (C, D) ordered according to their expression level and the y-axis represents the abundance as percentage of reads (%) (average of all samples). Error bar shows SD.

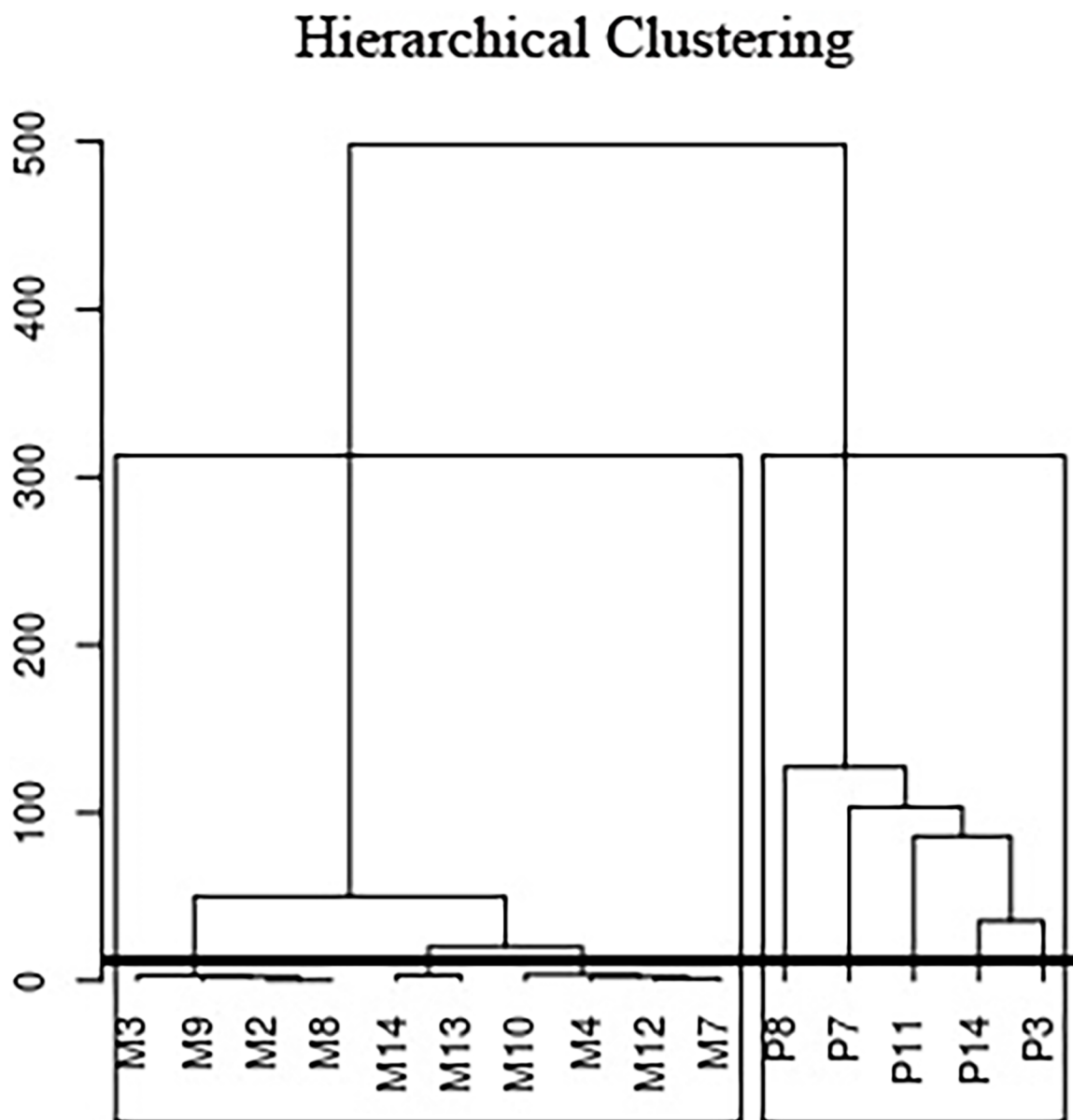
<https://doi.org/10.1371/journal.pone.0193527.g001>



To further confirm the results given the different sequencing depth of each biofluid, 10 random subsamples of 0.2 M miRNA reads were obtained for each plasma and milk sample. The total number of unique miRNA in the 10 milk subsamples was 534 (115 common to all milk samples) and in the 10 plasma subsamples was 715 (100 common to all plasma samples). Samples were again classified by biofluid (S8 Fig). Two-hundred and fourteen miRNAs exhibited different levels depending on the biofluid [72 (33.64%) milk > plasma]. One hundred and eighty six overlapped with the results of the main analysis (S7 Fig, S4 Table).

### isomiRs

Total number of isomiRs with >1 count was 10,541 in milk (1,798 with mean abundance of >10 RPM) and 4,760 in plasma (1,376 with mean abundance of >10 RPM) (S5 Table). Some



**Fig 2.** Dendrogram of samples according to their normalized miRNA levels (10 milk samples (M) and 5 plasma samples (P)). Samples are classified by biofluid and not by individual as indicated by the squares.

<https://doi.org/10.1371/journal.pone.0193527.g002>

**Table 3. Differential levels of miRNAs by biofluid—top 10 ordered by p value.**

<b>A. Top ten higher in plasma</b>				
miRNA	baseMean	log2FC	p value	adjusted p value
hsa-miR-451a	26988.57	11.57	5.09e-62	3.59e-59
hsa-miR-122-5p	117899.75	13.14	9.14e-47	3.23e-44
hsa-miR-486-5p	531306.29	14.15	1.26e-42	2.96e-40
hsa-miR-4508	790.10	10.21	5.36e-38	9.47e-36
hsa-miR-134-5p	500.38	9.47	6.17e-24	8.71e-22
hsa-miR-363-3p	451.88	6.03	2.53e-21	2.98e-19
hsa-miR-584-5p	319.49	7.77	8.87e-21	8.95e-19
hsa-miR-150-5p	815.24	6.25	2.00e-20	1.66e-18
hsa-miR-370-3p	317.26	9.52	2.11e-20	1.66e-18
hsa-miR-127-3p	670.08	7.22	8.02e-20	5.66e-18
<b>B. Top ten higher in milk</b>				
miRNA	baseMean	log2FC	p value	adjusted p value
hsa-miR-200b-3p	22349.79	-7.74	2.60e-19	1.67e-17
hsa-miR-30a-5p	90191.79	-5.59	1.34e-18	7.88e-17
hsa-miR-200a-3p	55145.39	-7.13	4.16e-18	2.26e-16
hsa-miR-200c-3p	31313.85	-7.34	7.23e-18	3.64e-16
hsa-miR-146b-5p	73258.87	-5.00	2.70e-17	1.27e-15
hsa-miR-200a-5p	2472.99	-8.19	4.87e-15	2.02e-13
hsa-miR-30b-5p	2335.32	-6.24	9.53e-14	3.54e-12
hsa-miR-141-3p	9204.94	-7.08	1.48e-13	5.24e-12
hsa-miR-193b-3p	797.36	-6.97	1.84e-13	6.18e-12
hsa-miR-30a-3p	5492.94	-5.51	2.22e-12	6.27e-11

Analysis was done with DESeq2. baseMean is the mean of normalized counts for all samples. log2FC is the log2 fold change between plasma and milk samples. Adjusted p value by the Benjamini-Hochberg method.

<https://doi.org/10.1371/journal.pone.0193527.t003>

miRNAs bear more than one modification type. In plasma, 3' additions, 3' trimming and 5' trimming modifications represented 31.77%, 48.20%, and 20.03% of all editions, respectively. In milk, distributions were rather similar (3' additions: 34.04%, 3' trimmings: 44.91%, and 5' trimmings: 21.05%). Table 4 describes the distribution and abundance by biofluid of different types of isomiRs, as combinations of 1, 2 or 3 modification types. The most common combinations in terms of number of isomiRs were: 3' addition + 3' trimming (27.98% in plasma and 34.39% in milk) and 3' trimmings (27.79% in plasma and 18.35% in milk). Regarding relative

**Table 4. Number and relative abundance of different types of isomiRs, as combinations of different editions, in milk and plasma.**

isomiR with different modifications:	Number of isomiRs <sup>a</sup>				Relative abundance (%) <sup>b</sup>			
	Plasma (N = 4760)		Milk (N = 10541)		Plasma		Milk	
	N	%	N	%	Mean	SD	Mean	SD
3' addition	549	11.53%	987	9.36%	21.75%	7.86	6.41%	1.24
3' addition + 3' trimming	1332	27.98%	3625	34.39%	8.11%	2.42	19.33%	6.98
3' addition + 3' trimming + 5' trimming	406	8.53%	1358	12.88%	0.35%	0.20	0.95%	0.35
3' addition + 5' trimming	181	3.80%	489	4.64%	0.24%	0.07	0.26%	0.11
3' trimming	1323	27.79%	1934	18.35%	66.89%	8.47	70.66%	8.46
3' trimming + 5' trimming	683	14.35%	1604	15.22%	1.76%	0.99	1.54%	0.15
5' trimming	286	6.01%	544	5.16%	0.90%	0.41	0.85%	0.18

<https://doi.org/10.1371/journal.pone.0193527.t004>



Table 5. Number of clusters and abundance in milk and plasma.

Cluster class	Number of clusters				Relative abundance (%) <sup>a</sup>			
	Plasma (N = 819)		Milk (N = 1,003)		Plasma		Milk	
	N	%	N	%	Mean	SD	Mean	SD
miRNA	181	22.10	191	19.043	18.20	11.88	46.76	16.41
rRNA	5	0.61	6	0.60	0.01	0.01	0.01	0.00
tRNA	78	9.52	85	8.48	1.00	0.23	30.91	15.49
snoRNA	33	4.03	34	3.39	0.07	0.04	0.24	0.14
snRNA	3	0.37	4	0.40	0.03	0.03	0.01	0.01
VTRNA	3	0.37	3	0.30	0.01	0.01	0.01	0.01
lncRNA	14	1.71	19	1.89	0.36	0.20	0.03	0.01
piRNA	9	1.10	12	1.20	32.03	5.86	8.42	2.94
gene	259	31.62	421	41.97	6.25	2.80	0.93	0.32
repeat	108	13.19	97	9.67	40.57	36.66	10.02	3.44
not annotated	126	15.38	131	13.06	6.44	5.38	2.66	1.75

<sup>a</sup> Average of relative abundance of all samples (number of reads of each cluster class in a particular sample divided by total number of reads in that sample).

<https://doi.org/10.1371/journal.pone.0193527.t005>

abundance, isomiRs with 3' trimming changes accounted for around 70% of total reads in both biofluids. They were followed by isomiRs with 3' additions, in plasma, and by isomiRs with a combination of 3' trimmings + 3' additions, in milk.

At a 5% FDR, 1,790 isomiRs (variants of 298 different miRNAs) showed different levels between biofluids [1,266 (70.73%) milk > plasma] (S6 Table).

## Small RNA clusters

In total, one thousand and fifty three small RNA clusters were identified in the 10 milk and 5 plasma samples. One thousand and three of them were found in milk with >1 count (930 with mean abundance of >10 RPM), and 819 in plasma (648 with mean abundance of >10 RPM). Only for descriptive purposes, clusters with multiple annotations were classified to one single RNA class following this prioritization: miRNA, rRNA, tRNA, snoRNA, small nuclear RNA (snRNA), vault RNAs (VTRNA), long non-coding RNA (lncRNA), piRNA, gene, and repeats. The number and proportion of clusters in each RNA class by biofluid can be found in Table 5. In both biofluids, more than 30% of the clusters identified were annotated to genes, and around 20% to miRNAs. Milk samples showed a higher proportion of reads mapping to miRNAs (46.76%) and tRNAs (30.91%), whereas plasma had a higher proportion of reads in clusters represented by piRNAs (32.03%), genes (6.25%), and repeats (40.57%) (Table 5). miRNA clusters in plasma corresponded to 18.20% of the total number of reads. Around 6% and 3% of the reads remained not annotated in plasma and milk, respectively.

Similarly to the miRNA analysis, the 10 most abundant clusters represented >75% of the reads in each biofluid (Fig 1, S7 Table). In milk, a tRNA cluster (*tRNA-Gly-GCC*, *tRNA-Gly-CCC2*, *COLQ*, *VAC14*, and several piRNAs) accounted for 21.67% of the reads, followed by *MIR148A* with 20.69%. In plasma, 37.94% were from a cluster that mapped to several piRNAs and *RXR* gene. The second most abundant cluster, with 21.04% of the reads, included *MIR486-1*, *MIR486-2*, and *ANK1* gene. Abundance profile along precursor of these four clusters is shown in S9 Fig.

The levels of 774 clusters were statistically significant different between plasma and milk [291 (37.60%) milk > plasma] (Table 6 and S8 Table). Besides differently detected miRNAs,

**Table 6. Differential levels of small RNA clusters by biofluid—top 10 ordered by p value.**

A. Top ten higher in plasma						
Cluster id	Cluster class	annotation	baseMean	log2FC	p value	adjusted p value
141	miRNA	<i>MIR451B, MIR451A<sup>a</sup></i>	27640.81	11.76	1.44e-48	1.53e-45
695	miRNA	<i>MIR122<sup>a</sup>, MIR3591</i>	114740.42	13.24	3.05e-36	1.61e-33
815	miRNA	<i>MIR486-1<sup>a</sup>, MIR486-2<sup>a</sup>, ANK1</i>	581529.14	14.32	6.79e-35	2.39e-32
114	miRNA	<i>MIR150<sup>a</sup></i>	878.09	6.47	3.93e-18	1.04e-15
362	gene	<i>TSHZ2, repeat</i>	1745.85	8.87	1.41e-17	2.99e-15
671	miRNA	<i>MIR127<sup>a</sup>, RTL1</i>	607.44	7.54	1.43e-16	2.51e-14
390	miRNA	<i>MIR363<sup>a</sup></i>	434.99	6.20	4.46e-15	5.89e-13
30	repeat	<i>Repeat</i>	4220.07	8.31	7.15e-15	8.40e-13
314	gene	<i>ARHGEF38, repeat</i>	3069.65	8.22	9.92e-15	1.05e-12
361	not annotated	<i>not annotated</i>	1429.83	7.93	1.39e-14	1.23e-12
B. Top ten higher in milk						
Cluster id	Cluster class	annotation	baseMean	log2FC	p value	adjusted p value
128	miRNA	<i>MIR200B<sup>a</sup></i>	23506.88	-7.91	8.64e-16	1.30e-13
127	miRNA	<i>MIR200A<sup>a</sup></i>	59945.49	-7.31	1.22e-14	1.17e-12
130	miRNA	<i>MIR200C<sup>a</sup>, LOC105369635</i>	31561.40	-7.42	2.74e-14	2.23e-12
963	miRNA	<i>MIR30A<sup>a</sup></i>	98503.77	-5.61	1.19e-13	7.88e-12
943	miRNA	<i>MIR146B<sup>a</sup></i>	76866.79	-5.01	1.40e-12	6.18e-11
659	miRNA	<i>MIR335, MeST</i>	2261.74	-6.84	2.24e-12	9.47e-11
825	tRNA	<i>tRNA-Trp-CCA-2-1</i>	1432.83	-6.70	1.64e-11	5.98e-10
794	tRNA	<i>tRNA-Asp-GTC-1-1, LOC643770</i>	2532.81	-8.52	5.14e-11	1.65e-09
131	miRNA	<i>MIR141<sup>a</sup>, LOC105369635</i>	9289.10	-7.18	7.28e-11	2.20e-09
849	tRNA	<i>tRNA-Val-TAC-1</i>	2848.07	-5.87	1.58e-10	4.19e-09

Analysis was done with DESeq2. baseMean is the mean of normalized counts for all samples. log2FC is the log2 fold change between plasma and milk samples. Adjusted p value by the Benjamini-Hochberg method

<sup>a</sup> Detected among the top 10 findings in the miRNA analysis.

<https://doi.org/10.1371/journal.pone.0193527.t006>

clusters of other RNA classes were identified. They included clusters mapped to genes (*TSHZ2*, plasma > milk), tRNAs (*tRNA-Trp-CCA-2-1*, milk > plasma), VTRNs (*VTRNA2-1*, plasma > milk), snoRNAs (*SNORD95*, plasma > milk), or lncRNAs (*HOTAIR*, plasma > milk). Their profiles can be seen in [S10 Fig](#). When considering isomiRs or small RNA clusters instead of miRNAs, samples also grouped by biofluid ([S5](#) and [S6 Figs](#)).

## Discussion

In this study, we identified several species of small RNAs, including miRNAs, piRNAs, tRNAs, snRNAs and snoRNAs in both plasma and breast milk. Their profiles were biofluid specific. As the identification of circulating small RNAs depends on the quality of the biological material, the quantification method and the bioinformatic pipeline, we discuss below the technical aspects followed during this study that can potentially affect the results obtained.

## RNA extraction and sequencing

The RNA extraction and sequencing yielded data of enough quality for all milk samples and 5 out of the 15 plasmas. Forty % of reads in plasma and 89.58% in milk passed the quality control step, with only 3.97% and 35.72% mapping to miRNAs, respectively. RNA input for plasma

and milk small RNA libraries were 150 ng and 500 ng, respectively, which respond to the different RNA quantity obtained for each biofluid. RNA yield is thus a key factor for the quality of sequencing data.

In plasma, the lack of 18S and 28S peaks in the RNA patterns obtained with the Bioanalyzer suggested the absence of RNA cell contamination. Indeed, plasma samples were centrifuged twice in <3.5 h after blood collection and the spectrophotometer haemolysis values were below 0.2, except for one sample. Given the low RNA yield obtained per mL of plasma (around 40 ng/mL), the addition of carriers during RNA extraction would be recommended [37].

In contrast, milk samples presented some signal at 18S and 28S, independently of whether samples were centrifuged one or two times before storing. Milk samples also showed systematically higher RNA levels by mL of biofluid than plasma samples (around 50-times higher levels). Given the higher amount of RNA obtained in milk and the presence of ribosomal RNA, we suspect contamination with the cellular fraction during skim milk sample separation. However, other studies reported similar RNA yields for skim milk (around 4000 ng/mL) as ours, and higher amounts in the milk fat and cellular fractions [38]. Contamination with cellular RNA is one of the main issues in the use of circulating free small RNAs as biomarkers of disease [39]. Therefore, all the findings shown here, as well as in other publications, should be interpreted considering this potential limitation.

### Small RNA bioinformatic analysis

Currently, the analysis of miRNA sequencing data is quite straight forward. In contrast, differences of one or few nucleotides in isomiRs with respect to reference miRNAs and frequent multicopy non-miRNA small RNAs in the genome impose several limitations to their bioinformatics analysis. To quantify small RNAs, we used SeqCluster, a tool that groups RNAs in clusters based on their sequence similarity. RNA clusters are defined as unique transcriptions units with potentially the same molecular function given their similar sequence, regardless of their genomic location. For instance, tRNAs which are present in different copies in the genome are clustered together. The reference databases and approach used in this study differ from the pipeline followed in two of the most recent publications on circulating small RNAs: Freedman et al. 2016 (ExceRpt, based on sRNABench) [19] and Yeri et al. 2017 (sRNABench) [21]. In sRNABench, reads are first mapped to miRNA, and remaining reads are mapped to other small RNAs using ENSEMBL 75. For isomiRs characterization, 3' additions and 3' and 5' trimming editions were considered in the analysis, as there are strong evidences of their processing [40–42]. In contrast, isomiRs with substitutions were discarded due to difficulty of mapping them unambiguously.

### Small RNAs in plasma

Ten out of the 824 miRNAs detected in plasma represented >70% of total number of reads mapping to miRNAs. Other authors have also reported few miRNAs accounting for around 50% of the total number of sequencing reads [19, 43]. In agreement with Yeri et al. [21] hsa-miR-486-5p was the most abundant miRNA in plasma (representing approximately 40% of the miRNA reads). A bias towards this miRNA in samples prepared with the Illumina kit has been reported in plasma-derived exosomes [20]. Ten and 8 of the 10 top miRNAs identified in this study were detected among the top 50 positions in Freedman et al. [19] and in Yeri et al. [21], respectively. The most abundant miRNA in Freedman et al was hsa-miR-451a, while in our sample set this miRNA was found in the 5<sup>th</sup> position. According to public data deposited in the miRmine database [44], hsa-miR-486-5p and hsa-miR-451a are the two most abundant miRNAs in plasma, and are produced by red blood cells [39]. Hsa-miR-122, which is highly

abundant in plasma, is mostly expressed in liver and enters easily into circulation, pointing towards its potential as biomarker for several liver pathologies [45].

In agreement with Freedman et al. [19] and Yeri et al. [21] we also found piRNAs, tRNAs, snoRNAs, snRNAs, VTRNAs and lncRNAs in plasma. The abundance of miRNAs, tRNAs, snoRNAs, snRNAs, and VTRNAs was in a similar range among previous studies. In contrast, there were strong discrepancies in the piRNA levels. Interestingly, a piRNA cluster mapping to several piRNAs and *RXRB* gene accounted for 37.94% of the reads in the plasma samples analysed in this study. After visual inspection of the genomic region, we realized that this cluster also overlaps with *Y RNA* genes (*RNAY4/RNAY4P10*). Indeed, Yeri et al. (2017) described a high proportion of sequences mapping to *Y RNA* in plasma (>60%) and a much lower abundance of piRNAs [21]. *Y RNAs* participate in chromosomal DNA replication and are involved in RNA stability and response to stress [46]. Circulating *Y RNAs* of 25–33 nt have been found in blood, within vesicles or as cell-free ribonucleoprotein (RNP) complexes [47–49]. Their function in biofluids, if any, is unknown and it cannot be discarded whether they are degradation products.

piRNAs were thought to be specific to germ cells, where they play essential roles in fertility and regeneration [50], but they have been described in plasma [19, 21], urine, and saliva [21] and in healthy and cancer somatic cells [51, 52]. In particular, Keam et al showed that Hiwi2 protein, one of the three human Piwi proteins, was ubiquitously expressed in somatic cells. The sequencing of the RNA co-immunoprecipitated with Hiwi2 protein identified a wide range of small RNA sequences of 18 to 34 nt that mapped to processed tRNA fragments, known piRNAs and coding genes, among others. The authors suggest that, in somatic cells, the Piwi-piRNA pathway participates in the regulation of protein translation [51]. Of note, the way how piRNAs are identified, through sequencing of Piwi bound RNAs, might introduce some noise in the piRNA sequence pool in reference databases.

Differences in small RNA abundance with previous publications [19, 21] can be explained by biological and technical factors. The quantification of small RNAs was performed using different NGS platforms, for which some bias to specific small RNAs have been described before [20]. Moreover, these studies used different bioinformatics pipelines. The small sample size of our dataset (5 plasma samples) in combination with the low and highly heterogeneous expression pattern of small RNAs in plasma with only a few of them detected in all samples might have biased our findings. Finally, our study included a highly specific population of post-partum mothers, while Freedman et al. and Yeri et al. enrolled participants from the general population and male athletes, respectively.

## Small RNAs in milk

Several studies have analysed miRNAs in different milk fractions (whole milk [53], skim milk [38, 54, 55], milk fat [38, 53, 55], or milk extracellular vesicles [56–58]), in different lactation points, and using different methods (NGS [53, 56, 58], qPCR [55, 57], or microarrays [54]). Long non-coding RNAs in milk microvesicles have also been reported [57].

In the present study focusing on skim milk, 10 miRNAs accounted for >70% of the reads mapped to miRNAs. Hsa-miR-148a-3p represented around 30% of the reads. This miRNA was previously reported to be an abundant miRNA in human milk fat [53] and milk exosomes/microvesicles [56, 58]. Of notice, this miRNA was only found in studies using NGS, but not those applying qPCR or microarrays. Hsa-miR-146b, hsa-miR-200c and hsa-30a-5p, among top miRNAs expressed in our sample set, were also reported to be highly abundant in milk fat, skim milk or milk exosomes/microvesicles [53, 55, 56, 58]. Moreover, some of them were also abundant in the milk cellular fractions [55]. A direct comparison of this study versus

previous publications is subjected to the differences in starting biological material, lactation period, and quantification methods.

Besides miRNAs, we detected other small RNAs, being tRNAs one of the most abundant class. tRNAs have been found before in several biofluids [19, 21], including milk exosomes [56, 58]. tRNA fragments seem to have distinct expression patterns and regulatory functions [59]. As far as we know, this is the first study describing the presence of piRNAs, snRNAs, snoRNAs and VTRNAs in milk. Whether breast milk tRNAs and other small RNAs can reach child circulation and modulate child physiology is of great interest and requires further investigation.

## Differential detection of RNAs in both biofluids

As expected, we observed a clear separation between biofluids in terms of their small RNA profile [21, 60]. Previous studies have shown that the miRNA profile observed in each biofluid recapitulates the miRNA profile of the main cell types releasing small RNAs in those biofluids (blood cells in plasma [39] and mammary epithelial cells in milk [55]).

Besides miRNAs, several small RNA clusters showed a different expression pattern between biofluids. As an example, levels of *VTRNA2-1* were higher in plasma than in milk. *VTRNA2-1* transcript is a putative tumour suppressor and modulator of innate immunity, which responds to environmental stressors during development through loss of imprinting [61]. *HOTAIR*, also detected at higher levels in plasma, is a long non-coding RNA that represses transcription of *HOXD* genes in trans. *HOTAIR* has been implicated in cancer [62]. We detected >10,000 isomiRs which beard one or several editions (3' addition, 3' trimming or 5' trimming): 4,760 in plasma and 10,541 in milk, processed from 707 and 875 miRNAs, respectively. Differentially isomiRs found in milk and plasma matched with the differentially expressed miRNAs detected in these two biofluids. Some studies suggest that specific isoforms of the same miRNA are the ones that have an effect on disease [63]. Further investigation is needed to corroborate whether isomiRs affected by specific diseases can be found in circulation, however our study shows a myriad of isomiRs which could be potentially used as biomarkers.

Although sequencing depth of milk and plasma samples was similar, the proportion of good quality reads mapping to miRNA was much higher in milk than in plasma. This could bias the results of the differential analysis (i.e. some miRNAs expressed at mid or low levels in milk might not have been detected in plasma due to the lower sequencing depth). DESeq2 normalizes using the geometric median of the ratio between the gene of each sample and the average among samples, but when there are important differences in library size, normalization methods might not to be sufficient [64]. To confirm the biofluid-specific findings, we applied a subsampling approach in the miRNAs analysis. This approach consists in a random selection of the same read number in all the samples of the study. To control for a potential selection bias, we performed ten random subsamplings. Approximately 60% (186 out of 308) of the miRNAs detected in the main differential analysis were confirmed in the subsampling analysis. miRNAs only detected in the main analysis and not in the subsampling analysis could be false positives resulting from the different depth of sequencing between biofluids.

## Summary

Plasma and milk samples showed a completely different small RNA profile. Several types of small RNA species were detected in both biofluids, including miRNAs, tRNAs, snRNAs, snoRNAs, lncRNAs, and piRNAs. The presence of different species of small RNAs in biofluids, besides miRNAs, opens the door to explore them as potential biomarkers of disease or as mediators of lactation health effects.

## Supporting information

### S1 Fig. Bioinformatic and statistical analysis pipeline.

(TIF)

### S2 Fig. Bioanalyzer RNA pattern for representative samples. Plasma (A) and milk (B—low cell contamination, C—potential cell contamination) samples.

(TIF)

### S3 Fig. miRNA complexity per sample. miRNA complexity is defined as the number of miRNA genes that are observed as a function of the number of miRNA reads. The x-axis represents the number of reads (the total number is indicated on the right). The gradient colour of the bar from white to black shows the incremental detection of distinct miRNA genes as more of the sequenced reads are considered.

(TIF)

### S4 Fig. Principal Component Analysis of normalized miRNA count data (10 milk and 5 plasma samples). A) Heat-map of the association between the first five principal components and the technical and biological variables. SAMPLE\_TYPE and STARTING\_RNA (RNA input for library preparation, which is related to SAMPLE\_TYPE) are strongly associated with the first principal component having both a p-value of association of $8.44 \times 10^{-8}$ . B) Accumulated variance explained by the first 10 principal components. In green, the bars corresponding to principal component one and two are highlighted. The sea-green line indicates the accumulated explained variance in each principal component. The dark-red dashed lines indicate the principal component which accumulated explained variance overload 90%, which corresponds to the ninth principal component with an explained variance of 90.82%. C) Scatter plot of the samples located on the two first principal components. Samples cluster in two groups corresponding to milk (red) and plasma (blue).

(TIF)

### S5 Fig. Principal Component Analysis of normalized isomiRs count data (10 milk and 5 plasma samples). A) Heat-map of the association between the first five principal components and the technical and biological variables. SAMPLE\_TYPE and STARTING\_RNA (RNA input for library preparation, which is related to SAMPLE\_TYPE) are strongly associated with the first principal component having both a p-value of association of 0.002. B) Accumulated variance explained by the first 10 principal components. In green, the bars corresponding to principal component one and two are highlighted. The sea-green line indicates the accumulated explained variance in each principal component. The dark-red dashed lines indicate the principal component which accumulated explained variance overload 90%, which corresponds to the tenth principal component with an explained variance of 92.86%. C) Scatter plot of the samples located on the two first principal components. Samples cluster in two groups corresponds to milk (red) and plasma (blue).

(TIF)

### S6 Fig. Principal Component Analysis of normalized small RNA clusters count data (10 milk and 5 plasma samples). A) Heat-map of the association between the first five principal components and the technical and biological variables. SAMPLE\_TYPE and STARTING\_RNA (RNA input for library preparation, which is related to SAMPLE\_TYPE) are highly associated with the first principal component having both a p-value of association of $1.52 \times 10^{-7}$ . B) Accumulated variance explained by the first 10 principal components. In green, the bars corresponding to principal component one and two are highlighted. The sea-green line indicates the accumulated explained variance in each principal component. The dark-red dashed lines



indicate the principal component which accumulated explained variance overload 90%, which corresponds to the seventh principal component with an explained variance of 91.04%. C) Scatter plot of the samples located on the first two principal components. Samples cluster in two groups corresponding to milk (red) and plasma (blue).

(TIF)

**S7 Fig. Venn diagrams of overlapped differentially expressed miRNAs according to different filtering strategies.**

(TIF)

**S8 Fig. Dendrogram of miRNA expression levels in milk and plasma subsamples.** For each sample (10 milk and 5 plasmas) 10 random subsamples of 0.2M reads of miRNA were obtained. Samples classify by biofluid and by sample of origin. Coloured branches represent 10 subsamples of the same sample of origin and bottom boxes the biofluid.

(TIF)

**S9 Fig. Abundance profile along precursor for main RNA clusters detected in milk and plasma.** Y-axis represents the abundance profile, and x-axis the positions in the precursor. A) Milk: *MIR148A*, B) Milk: tRNA-Gly-GCC and tRNA-Gly-CCC; C) Plasma: Several piRNAs and *RXR* gene; D) Plasma: *MIR486-1*, *MIR486-2*, and *ANK1* gene. In orange milk samples, and in blue plasma samples.

(TIF)

**S10 Fig. Abundance profile along precursor for a set of representative small RNA clusters differentially expressed in milk and plasma.** Y-axis represents the abundance profile, and x-axis the positions in the precursor. A) HOTAIR, higher levels in plasma; B) SNORD95, higher levels in plasma; C) tRNA-Trp-CCA-2-1, higher levels in milk; D) VTRNA2-1, higher levels in plasma. In orange milk samples, and in blue plasma samples.

(TIF)

**S1 Table. RNA concentration (ng/ml) and purity, by sample.**

(XLSX)

**S2 Table. Quality control and mapped reads [mean and (SD)] in different species of small RNAs, by sample.**

(XLSX)

**S3 Table. Relative abundance of miRNAs in plasma and milk.**

(XLSX)

**S4 Table. Differential detection of miRNAs by biofluid—ordered by p value.**

(XLSX)

**S5 Table. Relative abundance of isomiRs in plasma and milk.**

(XLS)

**S6 Table. Differential detection of isomiRs by biofluid—ordered by p value.**

(XLS)

**S7 Table. Relative abundance of small RNA clusters in plasma and milk.**

(XLSX)

**S8 Table. Differential detection of RNA clusters by biofluid—ordered by p value.**

(XLSX)

## Author Contributions

**Conceptualization:** Mariona Bustamante, Silvia Agramunt, Ramon Carreras, Xavier Estivill, Juan R. Gonzalez, Alfredo Mayor.

**Data curation:** Carles Hernandez-Ferrer.

**Formal analysis:** Mercedes Rubio, Carles Hernandez-Ferrer, Dietmar Fernandez-Orth, Juan R. Gonzalez.

**Funding acquisition:** Xavier Estivill, Juan R. Gonzalez, Alfredo Mayor.

**Investigation:** Mercedes Rubio, Mariona Bustamante, Dietmar Fernandez-Orth, Yaris Sarria, Maria Piqué-Borras.

**Methodology:** Mercedes Rubio, Mariona Bustamante, Carles Hernandez-Ferrer, Dietmar Fernandez-Orth, Yaris Sarria, Maria Piqué-Borras.

**Project administration:** Mercedes Rubio, Mariona Bustamante, Alfredo Mayor.

**Resources:** Yaris Sarria, Maria Piqué-Borras, Kilian Vellve.

**Supervision:** Mercedes Rubio, Mariona Bustamante, Lorena Pantano, Kilian Vellve, Silvia Agramunt, Ramon Carreras, Alfredo Mayor.

**Validation:** Lorena Pantano, Juan R. Gonzalez, Alfredo Mayor.

**Writing – original draft:** Mercedes Rubio, Mariona Bustamante, Carles Hernandez-Ferrer, Dietmar Fernandez-Orth, Lorena Pantano, Juan R. Gonzalez, Alfredo Mayor.

**Writing – review & editing:** Mercedes Rubio, Mariona Bustamante, Carles Hernandez-Ferrer, Dietmar Fernandez-Orth, Lorena Pantano, Yaris Sarria, Maria Piqué-Borras, Kilian Vellve, Silvia Agramunt, Ramon Carreras, Xavier Estivill, Juan R. Gonzalez, Alfredo Mayor.

## References

1. Strimbu K, Tavel JA. What are biomarkers? *Current Opinion in HIV and AIDS*. 2010; 5(6):463–6. <https://doi.org/10.1097/COH.0b013e32833ed177> PMID: 20978388
2. Siddeek B, Inoubli L, Lakhdari N, Rachel PB, Fussell KC, Schneider S, et al. MicroRNAs as potential biomarkers in diseases and toxicology. *Mutat Res Genet Toxicol Environ Mutagen*. 2014; 764–765:46–57. Epub 2014/02/04. <https://doi.org/10.1016/j.mrgentox.2014.01.010> [pii]. PMID: 24486656.
3. Xie T, Huang M, Wang Y, Wang L, Chen C, Chu X. MicroRNAs as Regulators, Biomarkers and Therapeutic Targets in the Drug Resistance of Colorectal Cancer. *Cell Physiol Biochem*. 2016; 40(1–2):62–76. Epub 2016/11/15. 000452525 [pii] <https://doi.org/10.1159/000452525> PMID: 27842308.
4. Pua HH, Ansel KM. MicroRNA regulation of allergic inflammation and asthma. *Curr Opin Immunol*. 2015; 36:101–8. Epub 2015/08/09. <https://doi.org/10.1016/j.coi.2015.07.006> S0952-7915(15)00103-X [pii]. PMID: 26253882; PubMed Central PMCID: PMC4593751.
5. Omran A, Elimam D, Yin F. MicroRNAs: New Insights into Chronic Childhood Diseases. *BioMed Research International*. 2013; 2013:13. <https://doi.org/10.1155/2013/291826> PMID: 23878802
6. Isaacs SR, Wang J, Kim KW, Yin C, Zhou L, Mi QS, et al. MicroRNAs in Type 1 Diabetes: Complex Interregulation of the Immune System,  $\beta$  Cell Function and Viral Infections. *Current Diabetes Reports*. 2016; 16(12):133. <https://doi.org/10.1007/s11892-016-0819-2> PMID: 27844276
7. Bilsland AE, Revie J, Keith WN. MicroRNA and Senescence: The Senectome, Integration and Distributed Control. 2013; 18(4):373–90. <https://doi.org/10.1615/CritRevOncog.2013007197> PMID: 23614622
8. Follert P, Cremer H, Beclin C. MicroRNAs in brain development and function: a matter of flexibility and stability. *Frontiers in Molecular Neuroscience*. 2014; 7(5). <https://doi.org/10.3389/fnmol.2014.00005> PMID: 24570654
9. Creemers EE, Tijssen AJ, Pinto YM. Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease? *Circ Res*. 2012; 110(3):483–95. Epub 2012/02/04. <https://doi.org/10.1161/CIRCRESAHA.111.247452> 110/3/483 [pii]. PMID: 22302755.

10. Vrijens K, Bollati V, Nawrot TS. MicroRNAs as potential signatures of environmental exposure or effect: a systematic review. *Environ Health Perspect*. 2015; 123(5):399–411. Epub 2015/01/24. <https://doi.org/10.1289/ehp.1408459> PMID: 25616258; PubMed Central PMCID: PMC4421768.
11. Hakimi MA, Cannella D. Apicomplexan parasites and subversion of the host cell microRNA pathway. *Trends Parasitol*. 2011; 27(11):481–6. Epub 2011/08/16. <https://doi.org/10.1016/j.pt.2011.07.001> S1471-4922(11)00122-X [pii]. PMID: 21840260.
12. Balzano F, Deiana M, Dei Giudici S, Oggiano A, Baralla A, Pasella S, et al. miRNA Stability in Frozen Plasma Samples. *Molecules*. 2015; 20(10):19030–40. Epub 2015/10/23. <https://doi.org/10.3390/molecules201019030> molecules201019030 [pii]. PMID: 26492230.
13. Ghai V, Wang K. Recent progress toward the use of circulating microRNAs as clinical biomarkers. *Arch Toxicol*. 2016; 90(12):2959–78. Epub 2016/09/03. <https://doi.org/10.1007/s00204-016-1828-2> [pii]. PMID: 27585665.
14. Amorim M, Salta S, Henrique R, Jeronimo C. Decoding the usefulness of non-coding RNAs as breast cancer markers. *J Transl Med*. 2016; 14:265. Epub 2016/09/16. <https://doi.org/10.1186/s12967-016-1025-3> [pii]. PMID: 27629831; PubMed Central PMCID: PMC5024523.
15. Witwer KW. Circulating microRNA biomarker studies: pitfalls and potential solutions. *Clin Chem*. 2015; 61(1):56–63. Epub 2014/11/14. <https://doi.org/10.1373/clinchem.2014.221341> clinchem.2014.221341 [pii]. PMID: 25391989.
16. Tan GC, Dibb N. IsomiRs have functional importance. *Malays J Pathol*. 2015; 37(2):73–81. Epub 2015/08/19. PMID: 26277662.
17. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011; 12(12):861–74. Epub 2011/11/19. <https://doi.org/10.1038/nrg3074> nrg3074 [pii]. PMID: 22094949.
18. Sobala A, Hutvagner G. Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. *RNA Biol*. 2013; 10(4):553–63. Epub 2013/04/09. <https://doi.org/10.4161/ma.24285> 24285 [pii]. PMID: 23563448; PubMed Central PMCID: PMC3710361.
19. Freedman JE, Gerstein M, Mick E, Rozowsky J, Levy D, Kitchen R, et al. Diverse human extracellular RNAs are widely detected in human plasma. *Nature communications*. 2016; 7:11106. <https://doi.org/10.1038/ncomms11106> PMID: 27112789; PubMed Central PMCID: PMC4853467.
20. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC genomics*. 2013; 14:319. <https://doi.org/10.1186/1471-2164-14-319> PMID: 23663360; PubMed Central PMCID: PMC3653748.
21. Yeri A, Courtright A, Reiman R, Carlson E, Beecroft T, Janss A, et al. Total Extracellular Small RNA Profiles from Plasma, Saliva, and Urine of Healthy Subjects. *Scientific reports*. 2017; 7:44061. Epub 2017/03/18. <https://doi.org/10.1038/srep44061> PMID: 28303895; PubMed Central PMCID: PMCPmc5356006.
22. Zempleni J, Aguilar-Lozano A, Sadri M, Sukreet S, Manca S, Wu D, et al. Biological Activities of Extracellular Vesicles and Their Cargos from Bovine and Human Milk in Humans and Implications for Infants. *J Nutr*. 2016. Epub 2016/11/18. jn238949 [pii] jn.116.238949 [pii] <https://doi.org/10.3945/jn.116.238949> PMID: 27852870.
23. Bagci C, Allmer J. One Step Forward, Two Steps Back; Xeno-MicroRNAs Reported in Breast Milk Are Artifacts. *PLoS One*. 2016; 11(1):e0145065. Epub 2016/01/30. <https://doi.org/10.1371/journal.pone.0145065> PONE-D-15-03369 [pii]. PMID: 26824347; PubMed Central PMCID: PMC4732600.
24. Alsaweed M, Hartmann PE, Geddes DT, Kakulas F. MicroRNAs in Breastmilk and the Lactating Breast: Potential Immunoprotectors and Developmental Regulators for the Infant and the Mother. *International journal of environmental research and public health*. 2015; 12(11):13981–4020. <https://doi.org/10.3390/ijerph121113981> PMID: 26529003; PubMed Central PMCID: PMC4661628.
25. Pantano L, Estivill X, Marti E. A non-biased framework for the annotation and classification of the non-miRNA small RNA transcriptome. *Bioinformatics*. 2011; 27(22):3202–3. Epub 2011/10/07. <https://doi.org/10.1093/bioinformatics/btr527> btr527 [pii]. PMID: 21976421.
26. Pantano L, Friedlander MR, Escaramis G, Lizano E, Pallares-Albanell J, Ferrer I, et al. Specific small-RNA signatures in the amygdala at premotor and motor stages of Parkinson's disease revealed by deep sequencing analysis. *Bioinformatics*. 2016; 32(5):673–81. Epub 2015/11/05. <https://doi.org/10.1093/bioinformatics/btv632> btv632 [pii]. PMID: 26530722.
27. Shah JS, Soon PS, Marsh DJ. Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs. *PLoS One*. 2016; 11(4):e0153200. Epub 2016/04/08. <https://doi.org/10.1371/journal.pone.0153200> PONE-D-15-53708 [pii]. PMID: 27054342; PubMed Central PMCID: PMC4824492.
28. Pantano L, Estivill X, Marti E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res*.

- 2010; 38(5):e34. Epub 2009/12/17. <https://doi.org/10.1093/nar/gkp1127> gkp1127 [pii]. PMID: 20008100; PubMed Central PMCID: PMC2836562.
29. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 2009; 10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: 19261174; PubMed Central PMCID: PMC2690996.
30. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011; 39(Database issue):D152–7. <https://doi.org/10.1093/nar/gkq1027> PMID: 21037258; PubMed Central PMCID: PMC3013655.
31. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004; 32(Database issue):D493–6. <https://doi.org/10.1093/nar/gkh103> PMID: 14681465; PubMed Central PMCID: PMC308837.
32. Zhang P, Si X, Skogerbo G, Wang J, Cui D, Li Y, et al. piRBase: a web resource assisting piRNA functional study. *Database: the journal of biological databases and curation*. 2014; 2014:bau110. <https://doi.org/10.1093/database/bau110> PMID: 25425034; PubMed Central PMCID: PMC4243270.
33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014; 15(12):550. Epub 2014/12/18. s13059-014-0550-8 [pii] <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281; PubMed Central PMCID: PMC4302049.
34. Lê S, Josse J, Hussen F. FactoMineR: An R Package for Multivariate Analysis. 2008; 25(1):18. Epub 2008-03-18. <https://doi.org/10.18637/jss.v025.i01>
35. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001; 125(1–2):279–84. Epub 2001/10/30. S0166-4328(01)00297-2 [pii]. PMID: 11682119.
36. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 1986; 73(3):751–4. <https://doi.org/10.1093/biomet/73.3.751>
37. Duy J, Koehler JW, Honko AN, Minogue TD. Optimized microRNA purification from TRIzol-treated plasma. *BMC genomics*. 2015; 16:95. <https://doi.org/10.1186/s12864-015-1299-5> PMID: 25765146; PubMed Central PMCID: PMC4342875.
38. Alsaweed M, Hepworth AR, Lefevre C, Hartmann PE, Geddes DT, Hassiotou F. Human Milk MicroRNA and Total RNA Differ Depending on Milk Fractionation. *Journal of cellular biochemistry*. 2015; 116(10):2397–407. Epub 2015/05/01. <https://doi.org/10.1002/jcb.25207> PMID: 25925799; PubMed Central PMCID: PMC45042114.
39. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyajima MM, et al. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer prevention research (Philadelphia, Pa)*. 2012; 5(3):492–7. Epub 2011/12/14. <https://doi.org/10.1158/1940-6207.capr-11-0370> PMID: 22158052; PubMed Central PMCID: PMC4186243.
40. Fernandez-Valverde SL, Taft RJ, Mattick JS. Dynamic isomiR regulation in Drosophila development. *RNA (New York, NY)*. 2010; 16(10):1881–8. Epub 2010/09/02. <https://doi.org/10.1261/ma.2379610> PMID: 20805289; PubMed Central PMCID: PMC42941097.
41. Vickers KC, Sethupathy P, Baran-Gale J, Remaley AT. Complexity of microRNA function and the role of isomiRs in lipid homeostasis. *Journal of lipid research*. 2013; 54(5):1182–91. Epub 2013/03/19. <https://doi.org/10.1194/jlr.R034801> PMID: 23505317; PubMed Central PMCID: PMC4222316.
42. Zhang J, Zhang S, Li S, Han S, Wu T, Li X, et al. A genome-wide survey of microRNA truncation and 3' nucleotide addition events in larch (*Larix leptolepis*). *Planta*. 2013; 237(4):1047–56. Epub 2012/12/13. <https://doi.org/10.1007/s00425-012-1823-7> PMID: 23232766.
43. Tonge DP, Gant TW. What is normal? Next generation sequencing-driven analysis of the human circulating miRNAome. *BMC molecular biology*. 2016; 17:4. <https://doi.org/10.1186/s12867-016-0057-9> PMID: 26860190; PubMed Central PMCID: PMC4748454.
44. Panwar B, Omenn GS, Guan Y. miRmine: A Database of Human miRNA Expression Profiles. *Bioinformatics*. 2017. <https://doi.org/10.1093/bioinformatics/btx019> PMID: 28108447.
45. Thakral S, Ghoshal K. miR-122 is a unique molecule with great potential in diagnosis, prognosis of liver disease, and therapy both as miRNA mimic and antimir. *Current gene therapy*. 2015; 15(2):142–50. PMID: 25537773; PubMed Central PMCID: PMC4439190.
46. Kowalski MP, Krude T. Functional roles of non-coding Y RNAs. *The international journal of biochemistry & cell biology*. 2015; 66:20–9. <https://doi.org/10.1016/j.biocel.2015.07.003> PMID: 26159929; PubMed Central PMCID: PMC4726728.
47. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Mote P, Martin DI. 5'-YRNA fragments derived by processing of transcripts from specific YRNA genes and pseudogenes are abundant in human serum and plasma. *Physiological genomics*. 2013; 45(21):990–8. <https://doi.org/10.1152/physiolgenomics.00129.2013> PMID: 24022222.

48. Dhahbi JM. Circulating small noncoding RNAs as biomarkers of aging. *Ageing research reviews*. 2014; 17:86–98. <https://doi.org/10.1016/j.arr.2014.02.005> PMID: 24607831.
49. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. *Biomarkers in cancer*. 2014; 6:37–47. <https://doi.org/10.4137/BIC.S20764> PMID: 25520563; PubMed Central PMCID: PMC4260766.
50. Clark JP, Lau NC. Piwi Proteins and piRNAs step onto the systems biology stage. *Advances in experimental medicine and biology*. 2014; 825:159–97. [https://doi.org/10.1007/978-1-4939-1221-6\\_5](https://doi.org/10.1007/978-1-4939-1221-6_5) PMID: 25201106; PubMed Central PMCID: PMC4248790.
51. Keam SP, Young PE, McCorkindale AL, Dang TH, Clancy JL, Humphreys DT, et al. The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. *Nucleic Acids Res*. 2014; 42(14):8984–95. Epub 2014/07/20. <https://doi.org/10.1093/nar/gku620> PMID: 25038252; PubMed Central PMCID: PMC4132735.
52. Spornraft M, Kirchner B, Pfaffl MW, Riedmaier I. Comparison of the miRNome and piRNome of bovine blood and plasma by small RNA sequencing. *Biotechnology letters*. 2015; 37(6):1165–76. Epub 2015/02/24. <https://doi.org/10.1007/s10529-015-1788-2> PMID: 25700822.
53. Munch EM, Harris RA, Mohammad M, Benham AL, Pejerrey SM, Showalter L, et al. Transcriptome profiling of microRNA by Next-Gen deep sequencing reveals known and novel miRNA species in the lipid fraction of human breast milk. *PLoS One*. 2013; 8(2):e50564. <https://doi.org/10.1371/journal.pone.0050564> PMID: 23418415; PubMed Central PMCID: PMC3572105.
54. Kosaka N, Izumi H, Sekine K, Ochiya T. microRNA as a new immune-regulatory agent in breast milk. *Silence*. 2010; 1(1):7. <https://doi.org/10.1186/1758-907X-1-7> PMID: 20226005
55. Alsaweed M, Lai CT, Hartmann PE, Geddes DT, Kakulas F. Human milk miRNAs primarily originate from the mammary gland resulting in unique miRNA profiles of fractionated milk. *Scientific reports*. 2016; 6:20680. <https://doi.org/10.1038/srep20680> PMID: 26854194; PubMed Central PMCID: PMC4745068.
56. Zhou Q, Li M, Wang X, Li Q, Wang T, Zhu Q, et al. Immune-related microRNAs are abundant in breast milk exosomes. *International journal of biological sciences*. 2012; 8(1):118–23. Epub 2012/01/03. PMID: 22211110; PubMed Central PMCID: PMC43248653.
57. Karlsson O, Rodosthenous RS, Jara C, Brennan KJ, Wright RO, Baccarelli AA, et al. Detection of long non-coding RNAs in human breastmilk extracellular vesicles: Implications for early child development. *Epigenetics*. 2016; 0. Epub 2016/08/06. <https://doi.org/10.1080/15592294.2016.1216285> PMID: 27494402; PubMed Central PMCID: PMC45094628.
58. Simpson MR, Brede G, Johansen J, Johnsen R, Storro O, Saetrom P, et al. Human Breast Milk miRNA, Maternal Probiotic Supplementation and Atopic Dermatitis in Offspring. *PLoS One*. 2015; 10(12):e0143496. Epub 2015/12/15. <https://doi.org/10.1371/journal.pone.0143496> PMID: 26657066; PubMed Central PMCID: PMC4682386.
59. Gebetsberger J, Polacek N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol*. 2013; 10(12):1798–806. <https://doi.org/10.4161/ma.27177> PMID: 24351723; PubMed Central PMCID: PMC3917982.
60. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. *Clin Chem*. 2010; 56(11):1733–41. <https://doi.org/10.1373/clinchem.2010.147405> PMID: 20847327; PubMed Central PMCID: PMC4846276.
61. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al. Independent genome-wide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome biology*. 2015; 16:118. <https://doi.org/10.1186/s13059-015-0660-y> PMID: 26062908; PubMed Central PMCID: PMC4464629.
62. Wang Y, Dang Y, Liu J, Ouyang X. The function of homeobox genes and lncRNAs in cancer. *Oncology letters*. 2016; 12(3):1635–41. <https://doi.org/10.3892/ol.2016.4901> PMID: 27588114; PubMed Central PMCID: PMC4998071.
63. Wang S, Xu Y, Li M, Tu J, Lu Z. Dysregulation of miRNA isoform level at 5' end in Alzheimer's disease. *Gene*. 2016; 584(2):167–72. <https://doi.org/10.1016/j.gene.2016.02.020> PMID: 26899870.
64. Tam S, Tsao MS, McPherson JD. Optimization of miRNA-seq data preprocessing. *Briefings in bioinformatics*. 2015; 16(6):950–63. <https://doi.org/10.1093/bib/bbv019> PMID: 25888698; PubMed Central PMCID: PMC4652620.